Considerations on modeling for early detection of abnormalities in locally autonomous distributed systems

Veelen, Martijn van

Chapter 4

Detection for Controlled Systems

Despite great efforts to model systems accurately, whenever a system is part of a physical reality, a model will not explain the systems behavior indefinitely. Wear, tear and unpredictable environments deform the system behavior continuously. Increasing deviation between models and reality reflects a decrease in the systems effectiveness. A sustaining effectiveness demands prevention of process failure and accommodation, and hence depends on early detection of disturbances. There are several strategies to disturbance detection founded on systems theory and statistical signal detection theory. In the wide variety of techniques, the cornerstone of detection is solving equations i.e. fitting parameters of a system or a statistical model from data. The effectiveness of strategies and techniques relates to the properties of the detection problem at hand. Therefore we classify the conventional detection arsenal to the properties of systems and abnormalities to which they apply. The anxious reader, with a background in systems and signal detection theory, just inspects the overview in section 4.5, and heads straight to chapter 5.

This chapter is a survey of conventional detection approaches for controlled systems. We start with an overview of key functions and base techniques, section 4.1. Then section 4.2 discusses proven methods from statistical detection theory, while section 4.3 similarly covers fault detection and isolation. This is complemented with a short discussion on contributions from computational intelligence and applications of neural networks for detection in section 4.4. In section 4.5 we provide a new perspective on the existing arsenal of techniques and strategies, relating the complexity of systems and abnormalities to suitable strategies. This novel classification narrows the scope of this research by hinting at the challenges following from the complexity of systems and abnormalities that may be expected in distributed systems.

4.1 Introduction

4.1.1 Background

Endeavours to model natural or man-made systems are sometimes inspired from shear curiosity but mostly meant to design or adapt a system towards a certain desirable behavior. Modeling starts from a fundamental understanding of the so-called first principles, i.e. the physical or logical laws dictating the behavior. The improvement of an existing model through observation is a key principle either to satisfy curiosity or to manipulate systems effectively. Differences between the presumed reality and observations occur. In pursuit of a better understanding or improved control over a process these differences must improve the existing models. The detection and isolation of differences precedes a possible diagnosis and accommodation of models. We take the perspective of systems to be designed or optimized, rather than of the pursuit of understanding as an objective in itself. In this perspective one or more objectives, e.g. production targets, are pursued, through desired behavior. We assume that if the behavior can

be influenced, i.e. processes exist that are controlled, then the system is controlled. Possibly the control is through an unknown organization i.e. an explicit controller may not yet have been identified. We also assume measurements of the behavior.

The goal of detection is to find and isolate deviations between desired and actual behavior leading to the failure to achieve the known objectives. Behavior is essentially formulated as models of signals and systems, i.e. the disciplines offering detection methods are Statistical Signal Detection (signals) and Fault Detection and Isolation (systems). The application domains of statistical signal detection (SDD) are communication and sensing. The application domains of FDI in design and other process operations [Venkatasubramanian, 2003] are:

1. Optimal sensor location: to enhance observability, detectability and separability;
2. Data reconciliation: detecting sensor faults, reconciliation of measurement;
3. Supervisory control: dealing with variability beyond blue print.

Our focus is on supervisory control with it's associated system requirements: availability, performance (e.g. the quality stability of product) and efficiency. The mechanisms for detection can come from either discipline, as they depend on the type of system, the type of abnormalities that can occur and the available knowledge. An overview of these mechanisms therefore starts from the different views on systems and abnormalities.

4.1.2 Views on systems and abnormalities

The designer's perception of the system, the disturbances (definition 2.7) and the resulting faults will determine the inception of the fault model and subsequently the optimal detection method for this model. The different types of faults and disturbances are in principle assumed to come from changes in fundamental parameters of the probability structure of state and input/output variables resulting in abnormalities, or in modifications in the probability structure itself which is not parameterized. The types of disturbances and associated fault models are

- static faults: those affecting only the static relationships or static parameters;
- dynamic faults: those affecting the dynamic relationship among variables.

These things bring the following first-hand observations:

- the dynamic faults are separated into switching and drifting (smooth) changes
- the signal detection theory has a model distinguishing the noise from the signal that comes as a (time) series of observations
- System theory provides the nominal model which uses the formalism of a stochastic differential/difference state space equation; fault models are embedded in this formalism to distinguish component, actuator, state and sensor faults and disturbances.

Information source and process model

The generic data generating process or information source has been introduced in section 2.1. Information sources are defined by three dimensions $I_\theta = (\chi, \Theta, (p_\theta)_{\theta \in \Theta})$. Abnormalities can be considered as deviations in either of these three dimensions. However a change in the random process can only be caused by a change in either Θ or $(p_\theta)_{\theta \in \Theta}$.

Types of disturbances

There are different types of faults and disturbances [Isermann, 1984], such as change in bias (drift); change in opposite direction (diffusion), change due to a certain pattern; increase in variance. Faults are changes in (causal) relationships linking significant variables [Rault & Baskiotis,1989], thus they can be static or dynamic in nature. Static errors are differences in the residue (marginal) distribution $p(\mathcal{E}_i|\theta)$ vs. $p(\mathcal{E}_i|\theta')$, or in the static relations (simultaneous) distribution of variables $p(\mathcal{E}_i, \mathcal{E}_i|\theta)$. vs. $p(\mathcal{E}_i, \mathcal{E}_i|\theta')$. Dynamic errors are spectral differences (differences in the dynamics of a variable) or the dynamic relation between variables in a multivariate stochastic process:

$$p(\mathcal{E}_i[n]|\mathcal{E}_i[n-k], \theta) \text{ vs. } p(\mathcal{E}_i[n]|\mathcal{E}_i[n-k], \theta') \tag{4.1}$$

$$p(\mathcal{E}[n]|\boldsymbol{X}[n-k], \theta) \text{ vs. } p(\mathcal{E}[n]|\boldsymbol{X}[n-k], \theta') \tag{4.2}$$

Dynamic errors can have a continuous smooth drifting behavior. The drifting behavior originates from a continuous differentiable state $\theta(t)$ for which $\dot{\theta} = \frac{d\theta}{dt}$ exists for every t. In drifting there is a transition from the initial state $\theta(t_0)$ to a different state $\theta(t_\infty)$. Dynamic errors can also result from switching between discrete states. Switching behavior originates from a series of states $(\theta_n)_{n \in N}$ of the source I_θ that are visited at intervals $[t_{n-1}, t_n]$ starting at $t = 0$. Note that the disturbance has structure and an error may be random, following definition 2.7 (see also figure 2.7).

We can easily describe the different manifestations of changes as errors or disturbances, but when is a disturbance a fault, or when will it lead to a fault? The term fault is generally defined as a departure from an acceptable range of an observed variable or a calculated parameter associated with a process [Himmelblau, 1978]. A departure from expected behavior of a process is an abnormality, as to distinguish harmful changes from faults.

Definition 4.1: abnormalities

Abnormalities are deviations in the manifest behavior of instances $I_1, I_2, \ldots$ of the information source I compared to the expected manifest behavior of that source, i.e. differences in Θ causing the dependencies among the measured variables $(V_t)_{t \in T}$ to differ

If a system has an intended purpose, which we assume in case of a controlled system, an abnormality is a fault if it causes the degradation of the effectiveness of a process fulfilling it's intended purpose [Isermann, 1984]. It is clear that the detection of a fault (or even an abnormality) requires at least a model of the expected or desirable behavior, as well as an error measure which quantitatively describes the effectiveness of a system. Another distinction can be made from the severity of a fault. In FDI literature these are known as levels of deterioration [Rault & Baskiotis,1989]: unsteady faults, steady faults + failure, catastrophic faults.

Fault Models

Fault models depict clearly the different views on signals and disturbances (or faults). One distinguishes between the signal model and the system model. The signal model [Hancock & Wintz, 1966] considers the observations $\boldsymbol{x}(t)$ to consist of a signal of interest $s(t)$ observed through a system with transfer $\boldsymbol{a}$ polluted with noise $\boldsymbol{n}(t)$.

$$\boldsymbol{x}(t) = \boldsymbol{a}s(t) + \boldsymbol{n}(t) \tag{4.3}$$

A system-oriented fault in FDI assumes a state space model, or a physically plausible model (section 2.3.4). One is refered to section 2.4.4 for the background on control system theory. A general formulation of this modeling approach [Trunov and Polycarpou, 2000] considers a linear state space model with modeling uncertainties n_x, n_y, time-profiles β_x, β_y and fault functions f_x, f_y, where the starting times are τ_x, τ_y.

$$\begin{aligned} \dot{x} &= Ax + \alpha(y, u, t) + n_z(x, u, t) + \beta_x(t - \tau_x)f_x(x, u) \\ y &= Cx + n_y(x, u, t) + \beta_y(t - \tau_y)f_y(u) \end{aligned} \tag{4.4}$$

A more common FDI model assumes linear relationships [Frank, 1990] everywhere. x is the $n \times 1$ state vector, u is the $p \times 1$ known input vector, y is the $q \times 1$ vector of measured outputs and A, B and C the nominal system matrices of known dimensions. The term Ed models the unknown inputs to the actuators and the dynamic process, Kf the actuator and component faults, Fd the unknown inputs to the sensors, and Gf the sensor faults. The time-evolutions of $f(t)$ and $d(t))$ are often unknown.

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Ed(t) + Kf(t) \\ y(t) &= Cx(t) + Fd(t) + Gf(t) \end{aligned} \tag{4.5}$$

These system models explicitly categorize faults in the component (CFD), the instrument/sensor (IFD), and the actuator (AFD) [Frank, 1990]. The caused faults and false alarms are [Frank, 1990]: 1. Actuator, mode or sensor faults; 2) modeling errors; 3) system noise and measurement noise. In system-oriented detection there is a key difference between gross parameter and structural changes [Venkatasubramanian, 2003]:

- **Gross parameter changes**. In any modeling, some processes may occur below the selected level of model detail. These processes are typically lumped as parameters include interactions across the system boundary. Parameter failures arise when there is a disturbance entering the process from the environment through one or more exogenous (independent) variables.
- **Structural changes** refer to changes in the process itself. They can be due to hard failures in equipment. Structural malfunctions result in a change in the information flow between various variables. To handle such a failure in a diagnostic system requires the removal of the appropriate model equations and the restructuring of the other equations to describe the current situation of the process.

Causality and evolution

Failures and faults are events, whereas abnormalities and disturbances are symptoms. Detection and diagnosis is possible because there are causal relationships between the root causes of events and the symptoms [Himmelblau, 1978]. The time-evolution in the behavior as a result of some changes (causes) in the information source is of crucial importance for the prevention and accommodation of faults. If a fault is to be prevented, a good model of the behavior is required to extrapolate the observed behavior. Diagnosis requires a known relation between symptoms and faults and accommodation requires a known relation between symptoms and causes. Knowledge of the causality and evolution is captured in models of the behavior. An outline of the detection process illustrates how this knowledge comes into play.

4.1.3 Process outline

Methodologies for detection are characterized by a mixture of: 1) The a priori process knowledge; 2). The search technique used [Venkatasubramanian, 2003]. A detector in either a signal detection or a FDI framework is conventionally designed in two steps: 1) parametric modeling; 2) likelihood testing [Patton, Frank and Clark, 1989]. This effectively means [Basseville, 1988]: 1. Transform to a stochastic change problem (generate residuals); 2. use statistical tools for solving and optimization of decision boundaries. Fundamentally detection is a form of hypothesis testing. The detector is a function that maps the observations into a decision space. This mapping is transforms data into a compact representation called signatures. In the signature computation knowledge of the systems and/or signals is used, which may rely on an explicit model of the system.

Comparing and hypothesis testing

Given input samples $x = \{x_i\}_{1 \le i \le n}$ and let F be the probability distribution x, then the form of the probability distribution F depends on some parameters $\theta = \{\theta_1, \ldots, \theta_m\}$, m not necessarily finite. For the general detection problem and given an observed sample ξ, there are two hypothese $H_0 : \theta = \theta^{(\text{normal})}$ and $H_1 : \theta = \theta^{(\text{abnormal})}$. In case the probability distribution is not known, the comparison can only be based on manifest behavior or non-parametric, i.e. samples $\{\xi_i|\theta\}_i$ of an instance I_θ. If manifest behavior $D = \{\xi_i\}_{i<t}$ of $I_{\Theta^{(\text{normal})}}$ is available, one compares:

$$P(\xi|\theta^{(\text{normal})}) \text{ against } P(\xi|\theta^{(\text{abnormal})}) \tag{4.6}$$

Detection is estimation

A detector is thus defined as a statistical mapping from measurements onto a decision space Γ. Any model of signals or systems becomes part of the mapping function. The decision space is partitioned into sections associated with each hypothesis, where at least the expected desirable behavior is defined, possibly only through a database of measurements associated with acceptable behavior.

Definition 4.2: detector

A detector is a mapping from observed samples $\xi = (x_n)_{n \in T}$ to a test statistic $\mathbf{d}(\xi) \in F^{(d)}$ for which at least one subspace $\Gamma^{(0)} \subset F^{(d)}$ is defined (associated with hypothesis H_I), that the observed sample does not differ structurally from whatever is expected:

$$d(\xi) \in \Gamma^{(0)} \subset F^{(d)} \Rightarrow H_I \tag{4.7}$$

In practice $d(\xi)$ is a variable w to be estimated from the data. The most simple test is $w > \tau$ for some threshold τ, while in parametric approaches $w = \hat{\theta}$ typically mean or variance. We

may say ‘detection is estimation’. Fitting data is a key function in detection. The techniques, the capabilities and limitations in estimation (section 2.4) will be considered in this chapter.

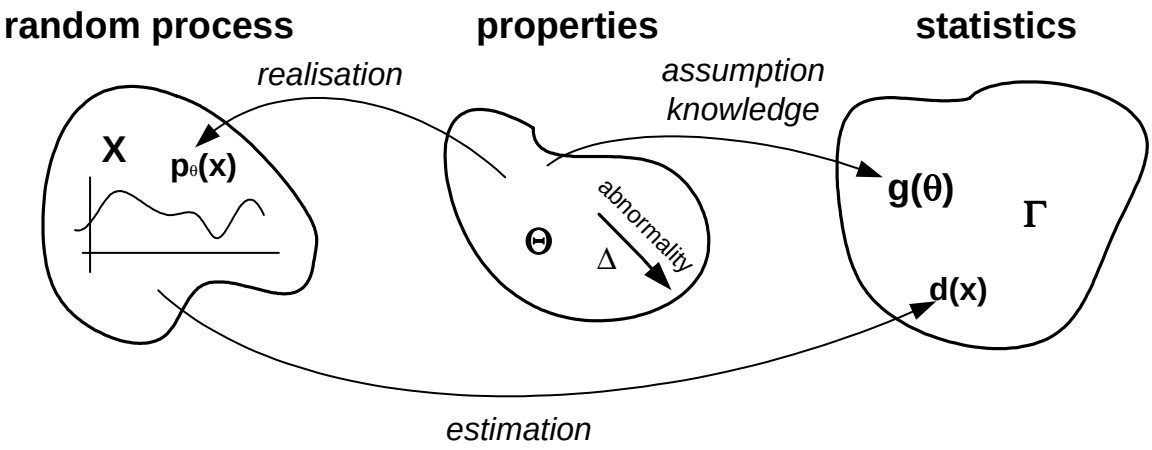

Figure 4.1 : Actual properties θ determine the possible realizations of a stochastic process. A change in these properties are abnormalities if they affect the performance of the system. Quantitative process history based methods aim to detect the change $\Delta\theta$ by estimating statistics from the data. See also figure 2.2.

Modeling of systems and faults

A priori knowledge comes from understanding the principles or laws of nature driving the information source. Together with an engineers preference and assumptions this is the a priori information. Selection of valid and reliable data precedes the modeling process and performance estimation of the detector. We define that model $\mathbf{M}$ represents the relationship between observed and controlled variables $(\boldsymbol{v}_n)_n$ (section 2.1.3) of the information source I_θ. Modeling has the goal of characterizing this structure by estimates of distributions or static and dynamic relations between observed variables $(\boldsymbol{v}_n)_n$. The better the model of systems and disturbances the better the detection. Therefore the techniques for modeling, introduced in section 2.3, will be considered for detection in this chapter.

Signature computation

The objective of residual generation is to isolate disturbances from random errors [Edwards et. al. 2000]. It can be decided from such a signature whether a fault has occurred [Patton, Frank and Clark, 1989]. The signature ϕ is a compact representation of the models behavior w.r.t. a specific sample ξ or database $\mathbf{D}$. Signature computation is either based on the residual $\mathcal{E}(\xi)$ or on the parameters of a model W. The statistical notion of sufficiency is important, particularly for signature computation where a sufficient statistic replaces the data for as far as the detector is concerned. Projection of data or parameter estimates or specific to detection, we consider statistical methods in section 4.2 and model-based FDI methods in section 4.3.

Design parameters

Some design parameters in the detection process are independent of the specific techniques applied for testing, modeling and signature computation. Examples are:

- **time-window** as selected for the signature computation. This parameter will determine the response time of the detector but also the reliability of the estimate. The time-window can be of fixed or variable size. A variable size is applied in sequential detection [Wald, 1946] for fixed detector confidence levels.

- **time-resolution** as mostly determined by the measurement system. However models of the system and it's faults can always have a lower time-resolution. The time-resolution may also not be fixed or non-equidistant sampling can be applied, see section 2.2.
- **number of samples** as used for detection is determined by time-resolution and the time-windows chosen for the off-line and on-line estimation.
- **detection thresholds** or **decision boundaries** as designed for the decision space
- the **number of learn-cycles** including the amount of data used to fit the initial model from the database as chosen for the off-line modeling of the existing data. The number of samples in the on-line estimation for detection is directly related to the re-estimation of model parameters, if re-estimation is applied in the detection approach.

The detection approach is firstly determined by the selected modeling approach and the signature computation. Secondly the quality of detection is to be optimized for all these design parameters. Both model selection as well as design optimization depends on the detection requirements and criteria that we will discuss hereafter.

4.1.4 Requirements and criteria

The requirements or objectives of detection can be derived is to improve the performance/ effectiveness and efficiency/availability of a system. Most of these objectives are hard to quantify without a specific system or fault model at hand. However, some basic generic criteria for detection have been defined.

Design objectives

The key objectives pursued in the detection are sensitivity, isolation, promptness and robustness. Sensitivity applies always to a priori known disturbances and faults, but in many applications detection is required to find abnormalities that are not or only partially specified.

Sensitivity to the known: Known faults and disturbances have to be detected and isolated from the observations. In particular they have to be separable in the decision space from the normal and acceptable behavior of the system.

Sensitivity to the unknown; novelty identifiability: Unknown and novel malfunction. One has access to a good dynamic model but it is possible that much of the abnormal operations region may not have been modelled adequately [Venkatasubramanian, 2003].

Isolation: Ability of the diagnostic system to distinguish between different failures. Under ideal conditions (free of noise and modeling uncertainties), a classifier should be able to generate output that is orthogonal to faults that have not occurred. This indicates a trade-off between isolability and the rejection of modeling uncertainties.[Venkatasubramanian, 2003].

Promptness: disturbances are to be detected before faults appear and before failure of the system as a whole. The less time is required for detection, the quicker the detector response.

Robustness: robust to various noise and uncertainties inclusive the performance to degrade gracefully instead of failing totally and abruptly [Venkatasubramanian, 2003]. Robust detectors are those designed to perform well, despite potentially damaging non-parametric deviations from a nominal parametric model [Dorf, 1993]. Robustness of a method is it's sensitivity w.r.t. it's design parameters [Isermann, 1984]. This depends on the complexity of the detection

algorithm and relies on the theoretical basis for the detection methodology. Therefore simpler methods are accepted better [Kitamura, 1989].

Reliability: An error estimate is often required for classification to build the user's confidence on reliability. Confidence levels on the diagnostic decisions should provide a priori estimates on the classification error that can occur. [Venkatasubramanian, 2003].

Adaptability: Process plants rarely remain invariant with periodic minor changes in operating policy, retrofit design and so on. Once a diagnostic system is deployed, it should be able to adapt with minimal effort as new situations are encountered and the scope of the system is expanded. [Venkatasubramanian, 2003]

Design criteria

Design criteria which actually provide a suitable numerical representation of a requirement are few. Well-known criteria are the False Detection Rate (FDR) and the False Acceptance Rate (FAR). Typically the null-hypothesis H_0 is the "normal operation" in detection. Then the FDR is the α defined in equation 4.8 and the FAR is the β defined in equation 4.9.

Type I error (FDR): $\alpha = p(d \in \Gamma^{(\text{abnormal})} \big| \theta \in \Theta^{(\text{normal})})$

$$\alpha = P(\boldsymbol{x} \in \Gamma_K \big| H) = \int_{\Gamma_K} f(\boldsymbol{x} \big| H) d\boldsymbol{x} \tag{4.8}$$

Type II error (FAR): $\beta = p(d \in \Gamma^{(\text{normal})} \big| \theta \in \Theta^{(\text{abnormal})})$

$$\beta = P(x \in \Gamma_H \big| K) = \int_{\Gamma_H} f(\boldsymbol{x} \big| K) d\boldsymbol{x} = 1 - \int_{\Gamma_K} f(\boldsymbol{x} \big| K) d\boldsymbol{x} \tag{4.9}$$

The probability of detection $1-\beta$ (also called the power of a test) is the probability of accepting an alternative when K is true, which means a correct detection of abnormality. The probability of detection is also a reasonable measure for the sensitivity. Reliability is quantified by the mean-time between false alarms (MTBF) [Basseville, 1988]. Reliability conflicts with sensitivity; an optimal balance between the two is mostly a cost-related matter. In some cases the expected cost can be derived algebraically or approximated given the probability of any type of disturbance. This is mostly not the case for novelty detection.

4.1.5 Key functions and base techniques

The detection principles are founded on two pillars: the conventional Statistical Signal Detection (SDD) theory and the Fault Detection and Isolation (FDI). The first is more data-driven while the second is model-driven. The key functions, as can be derived from the requirements and the process discussed above, are: 1) modeling of signals or systems and abnormalities; 2) estimation or fit; 3) testing, and 4) design/optimization of the models, estimations and the thresholds used in the hypothesis testing.

Modeling

There are different models for set-point control optimization and for detection:

- to specify the desired behavior by modeling systems and signals;

- to describe the elementary process (excluding controller): the nominal process model;
- to model hypothesized faults and disturbances;
- to simulation and extrapolation from behavioral system models.

The statistical modeling of signals and disturbances assumes a random process as an information source, modeling probability distributions through statistical dynamic models. These are called parametric models for known probability distributions, section 2.3.2 and 2.3.3. They are non-parametric in case one does not assume a probability distribution. In case the underlying principles (physics or logic) are known, system theory can be applied to obtain a physically plausible explanation (2.3.4). Often observers and controllers can be or have been designed by applying principles of control system theory.

Fitting

Fitting data is a key function in detection. Both for obtaining a model of the data in the database as well as for on-line parameter re-estimation or model identification. Essentially the estimation problem, section 2.4, is to solve a set of equations. The tools to do so from linear algebra and statistics are the Least Squares solution, Eigen Value Decomposition, Singular Value Decomposition, Partial Least Squares and Maximum likelihood. A summary of these tools is included in appendix C.1.

Testing

If information on the physical process is available to determine optimal parameter bounds through algebraic manipulation, then the actual test to choose an hypothesis relies on either threshold logic or parameter bounds. Often the hypothesis test is a simple test on mean or variance of the computed signatures, as one can assume a stationary residual in case of normal operation. Alternatively, normality and trend can be measured (subsection 2.2.2). In the absence of a parametric model it is possible to compare two samples directly without using a statistic. A brief overview of simple statistical tests is presented in subsections 4.2.2 and 4.2.3.

Design

In design one pursues optimization of some qualities by exploiting the relationship between detection criteria, data properties and design parameters. In detection there are two optimization issues: 1) for the models; and 2) for the tests. Either of these design issues has fundamental theoretical bounds. In the previous chapter we already discussed some procedure and measures for optimal model design, such as the optimal model order in relation to complexity of the problem (recall also the procedure of linearization and controller and observer design). We have also covered the fundamental bounds in statistical estimation (for instance, the Cramer-Rao bound), the limitations in linear solving (like singularities due to over- and under-determination), and recall also the observability and controllability matrices which can identify fundamental limitations to state identifiability. In addition to those there are some basic procedures relevant in the context of detection, like the Neyman-Pearson Strategy, Bayes optimal detector and the Likelihood Ratio Test.

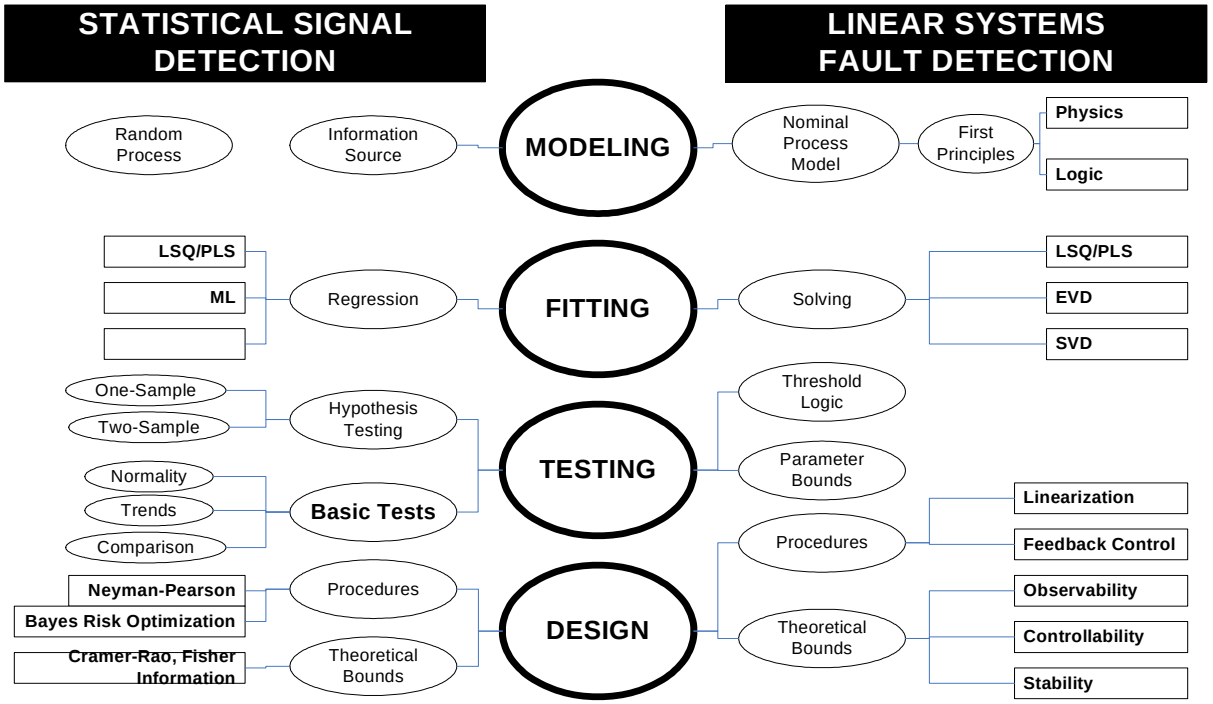

Figure 4.2 : An overview of the statistical and algebraic theory implementing key detection functions.

4.2 Statistical signal detection

4.2.1 Preliminaries

Statistical signal detection is applied for communication and sensing. The challenge is to detect and isolate a signal of interest from an observed signal containing noise and interfering signals. In the discussion on detection for controlled systems signal detection theory may appear the odd one out. However techniques from signal detection theory are applied in both model-free as well as model-based, e.g. system-oriented, detection. In quantitative detection one arrives at some point at a residual, i.e. the difference between expected behavior and actual observations. Key functions are hypothesis testing and search through (signature) projection of data. Design and optimization of decision function (like hypothesis tests) is well covered in signal detection theory. Data analysis, as discussed in section 2.2.2, is the basis of signature computation and model design. In the context of detection for controlled systems, an ideal stable operation generates random errors and no structural deviations. The detection of disturbances starts from the analysis of the residual system errors, a simple stationarity test (section 2.2). There are however different ways to project the measurements onto a decision space, starting from the choice of an hypothesis test.

Different hypothesis tests

We have included the basic examples of hypothesis testing in appendix C.1. There are a few basic design choices in the design of a detection test:

- **single-sample vs. two-sample**. In a single-sample test a property of the data is compared to some reference value, e.g. a threshold. Thus only a sample with new measurements is required of which some statistics are computed. In a two-sample test the two samples are either compared or a dependence between the two is estimated. The two-sample test is

applied when there is a database with reference samples corresponding to known types of behavior (like a sample of the acceptable behavior) to compare to new measurements.

- **single vs. two-sided**. In a single-sided hypothesis test, the null hypothesis is of the form $d(\xi) > \tau$; in a two-sided or symmetric test the null hypothesis is of the form $d(\xi) = \theta$ and the alternative is a deviation to either side of the reference value θ.
- **composite vs. simple**. In a simple hypothesis test the null hypothesis and it's alternative are complementary statements on the same property, e.g. on the mean of a sample: $H_0: \mu(\xi) = 0; H_1: \mu(\xi) \neq 0$. In a composite test the null hypothesis and the alternative are not complementary or on the same property, i.e. they can be heterogeneous.

Statistics

Statistics are compact representations of data, mostly for the purpose of analysis and testing. Important properties of statistics are bias, risk and sufficiency (see chapter 2 and appendix B.2). There is a key difference between parametric and non-parametric statistics. In parametric statistics a certain distribution is assumed such that the data can be represented by the distribution parameters. Other properties represented by statistics are non-parametric. They are useful only if they are robust for distributions associated with the decision regions (specific to each hypothesis), i.e. asymptotic non-parametric.

4.2.2 Basic one-sample tests: residual analysis

The most simple tests on residuals or other signals are one-sample tests without specific projections. The most straightforward tests are boundary tests; a bit more elaborate are some data-analysis methods valid for specific distributions, i.e. parametric one-sample tests. Finally some basic generic tests use a representation of the data which works for any distribution, e.g. histograms or frequency spectrums.

Boundary tests

The purpose of a boundary test is to check a limit or range. These limits are often determined from knowledge on the distribution of the data, conditional on either hypothesis. The simplest boundary test is given in equation 4.10. It is a boundary on any error in a residual ξ.

$$d(\xi) = \begin{cases} 1 \text{ if } \exists \boldsymbol{\nu} \in \xi : \left|e_i(\nu_i)\right| > \varepsilon_i \\ 0 \text{ otherwise} \end{cases} \tag{4.10}$$

The test can give a negative result already for a single error. Therefore it is entirely not robust to outliers, and it does not distinguish structural errors from incidental ones. Slightly more robust is the total power (or total received energy detector) equation 4.11. Typically a windowed version of these signatures is used e.g. $\phi_i[n] = \Sigma_{k=n-p}^{n} e_i[k]$. It responds to an increased average of the residual rather than a single outlier.

$$d(\xi) = \begin{cases} 1 \text{ if } \sum_{\boldsymbol{\nu} \in \xi} e_i(\nu_i)^2 > \varepsilon_i \\ 0 \text{ otherwise} \end{cases} \tag{4.11}$$

In case multiple sensor signals are available, the sensor signals observed are probably not independent, i.e. the mixing matrix $\boldsymbol{a}$ in the signal model is not a diagonal matrix. Rather than using an univariate threshold on each observed variable, a more robust test for zero-mean cross-correlated sensors is available [Wilsky, 1976]. Sensor errors are assumed to be normally distributed with a zero mean and cross-correlated with a known cross-correlation matrix C. This is limited to a linear dependence between the sensors and assumes absence of any dependence between the errors e_j in the vector $\boldsymbol{e}$. Ideally the dependencies are explained by a model, such that in case of normal operation the residual is independent.

$$\phi[k] = \sum_{i=k-N+1}^{\kappa} \boldsymbol{e}^{\mathrm{T}}[j] C^{-1} \boldsymbol{e}[j] \tag{4.12}$$

Parametric Tests

In case the residuals are multivariate and can be assumed i.i.d, the correlation coefficient can be used as a measure of structure. Statistics test for normality are the Lilifors Test [Sprent, 1984] and the Rao test, equation 4.13 [Wahlberg and Gustofsson, 2005].

$$L(x) = \frac{d}{d\theta}\log p(x|\theta)\Big|^{T}_{\theta_0|H_0} [I^{-1}(\theta_0|H_0)] \frac{d}{d\theta}\log p(x|\theta)\Big|_{\theta_0|H_0} \tag{4.13}$$

Variable sample size fixed confidence

In many detectors the time-windows, and hence the number of samples, is chosen to be fixed. However, particularly in non-parametric tests, the confidence in decisions often depends on the measurements themselves. Rather than by fixing the number of used samples, the confidence level of a test should be fixed, i.e. the detection delay resulting from the number of measurements should depend on a fixed FAR [Basseville, 1998] This is implemented by Wald's sequential test [Wald, 1946]. The optimal number of samples, or stopping time, is given by the Page-Hinkley stopping rule, valid for normal distributed errors [Basseville, 1998].

$$\begin{aligned} d((\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m)) &\in \Gamma^{(\text{abnormal})} \Rightarrow \text{accept } H_I \\ d((\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m)) &\in \Gamma^{(\text{normal})} \Rightarrow \text{accept } H_J \\ d((\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m)) &\in \Gamma^{(\text{indecisive})} \Rightarrow \text{take more data or increase m} \end{aligned} \tag{4.14}$$

Non-parametric tests and Structure Tests

A non-parametric efficient representation can be given by a histogram (equation 4.15). The histogram is an estimate for the probability density of the data. Another possible projection is the Fourier Transform. Detection can be based on a histogram by putting a threshold on each bin in a histogram or frequency spectrum.

$$H_i^{(\boldsymbol{h})}(X) = P(X \in [h_i, h_{i+1}]), \boldsymbol{H}^{(h)}(X) \equiv (H_i^{(\boldsymbol{h})}(X))_{0 \le i \le |h|-2} \tag{4.15}$$

The number of cells is $n_H = \#h - 1$ where h is a vector representing the cell boundaries, while Δh_x represents the cell width. Thresholding individual bins is not very robust. Alternatively the amount of structure in the data can be estimated. This is done by the information-theoretical measure of self-information. Self-information (equation 4.16) is the discrete estimation of the entropy and provides a good non-parametric test on the amount of structure present in data.

$$\left(\hat{\mathbf{H}}(x) = -\sum_i \left(\frac{H_i}{n_H}\log\left(\frac{H_i}{n_H}\right)\right) + \log(\Delta h_x)\right) \tag{4.16}$$

4.2.3 Basic two-sample tests for residual comparison

The two-sample tests are analog to single-sample tests categorized by straight comparison. Examples are parametric tests and robust parametric tests. Non-parametric tests also deal with un-equal sample size. Finally we mention structure tests. The generic result of statistical detection theory for two-sample tests is the Generalized Likelihood Ratio (GLR) tests.

Comparison

The most simple two-sample test is a straightforward comparison between two series of residuals, one serving as a reference signal associated with the hypothesized behavior. Such a comparison can be implemented using different norms and a threshold, e.g. the geometrical distance between the samples. This works only when the samples have an equal size. It is not very robust being sensitive to outliers and to time-shifts.

Parametric tests

The Generalized Likelihood Ratio Test (GLRT) can be computed given a parametric probability distribution $p_\theta(x)$ and hypothesis on the parameter(s) θ. For the likelihood of the sample under the hypothesis $H_0{:}\theta = \theta_1$ and $H_1{:}\theta = \theta_2$, the expressions are $p(\xi|\theta_1)$ and $p(\xi|\theta_2)$, resp. Their ratio l_θ, equation 4.17, leads to a decision in favor or against H_0.

$$l_\theta = \frac{p(\xi|\theta_2)}{p(\xi|\theta_1)} \tag{4.17}$$

A similar test, particularly suitable for normal distributions, is the Kullback Difference Measure (KDM). This is also a two-sample test, given the two probability density estimates of the samples: p_1 and p_2 resp. Then the measure and the discrete, multidimensional generalization, given the matrices with means $(M_1\ M_2)$ and the covariances $(\Sigma_1\ \Sigma_2)$ are

$$I_{1,2} = \int p_1(x)\left(\log\frac{p_1(x)}{p_2(x)}\right)dx \tag{4.18}$$

$$I_{1,2} = \frac{1}{2}\mathrm{tr}(\Sigma_2^{-1}\Sigma_1) + \frac{1}{2}(M_1 - M_2)^T\Sigma_2^{-1}(M_1 - M_2) - \frac{1}{2}\log(\det\Sigma_1/\det\Sigma_2) \tag{4.19}$$

The Jensen Difference Measure (GJDM) assumes two multinominal distributions of equal length $x = (x_1 \ldots x_2)$ and $y = (y_1 \ldots y_n)$. The Generalized Difference Measure with parameters $\pi = (\pi_1 \ldots \pi_2)$ is given in equation 4.21.

$$J_n(x,y) = H_n\left(\frac{x+y}{2}\right) - \frac{1}{2}(H_n(x) + H_n(y)) \tag{4.20}$$

$$J_n^\pi(y_1 \ldots y_k) = H_n\left(\sum_{i=1}^{k} \pi_i y_i\right) - \sum_{i=1}^{k} \pi_i H_n(y_i) \tag{4.21}$$

Robust parametric tests

In a similar fashion to Wilsky's multiple sensor tests, robust two-sample tests correct for inter-dependence between parameters, using the Fisher Information Matrix I for normalization. The Wald test [Wahlberg & Gustofsson, 2005] is given by

$$L(x) = (\hat{\theta} - \theta_0)^T [I^{-1}(\hat{\theta}_0 \big| H_1)](\hat{\theta} - \theta_0) \tag{4.22}$$

Non-parametric tests and unequal sample size

The aim is to test whether the samples result from a similar distribution without assuming which distribution they have. A comparison of the distributions is possible through a non-parametric estimation of the distributions by ordering the measurements into two samples. These orderings are $\hat{P}_1(x) = \text{sort}(\xi_1)$ and $\hat{P}_2(x) = \text{sort}(\xi_2)$. A histogram is a reasonable alternative estimate which also works for unequal sample lengths. The Cramer-Von Mises Statistics (CvM) is a two-sample test for samples of different sizes (N_1 and N_2 resp.) using the average square distance between the estimated distributions (equation 4.23). If the data is normally distributed, the test significance is 5% for a threshold T > 0.461, and the significance is 1% for the threshold T> 0.743.

$$T = N_1 N_2 \sum_i \frac{(P_1(x) - P_2(x))^2}{(N_1 + N_2)^2} \tag{4.23}$$

A more sensitive similarity measure is the maximum distance (equation 4.24). This is the Kolmogorov-Smirnov Test (KST) [Sprent, 1989] implementation [Press et al., 1992]; the confidence level (significance) depends on the effective number of observations N_e, while the actual test is given by equation 4.25 and 4.26.

$$D = \max_{\infty < x < \infty} \left| \hat{P}_1(x) - \hat{P}_2(x) \right| \tag{4.24}$$

$$Q_{\text{KS}}(\lambda) = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2\lambda^2} \tag{4.25}$$

$$P(D > \text{observed}) = Q_{\text{KS}}\left(\left[\sqrt{N_e} + 0.12 + \frac{0.11}{\sqrt{N_e}} \right] \cdot D \right) \text{ with } N_e = \frac{N_1 N_2}{N_1 + N_2} \tag{4.26}$$

Structure tests

The correlation between the two samples can be used as an estimate for their similarity. However this is limited to linear dependencies. Refraining from the use of a model is possible using information-theoretical measure such as the mutual entropy or the mutual information [Moddemeijer, 1989], using a multi-variable histogram.

$$\hat{\mathbf{H}}(x, y) = -\sum_{i,j} \left(\frac{H_{ij}}{n_H} \log\left(\frac{H_{i_j}}{n_H} \right) \right) + \log(\Delta h_x \Delta h_y) \tag{4.27}$$

$$\hat{\mathbf{I}}(x,y) = \sum_{j=0}^{J-1}\sum_{i=0}^{I-1}\frac{k_{ij}}{k_{\cdot\cdot}}\log\frac{k_{ij}\cdot k_{\cdot\cdot}}{k_{i\cdot}\cdot k_{\cdot j}} - \frac{(J-1)(I-1)}{2N} \tag{4.28}$$

The mutual information is alternatively computed $\mathbf{I}(x,y) = \mathbf{H}(x) + \mathbf{H}(y) - \mathbf{H}(x,y)$. This implies entropy can only be determined by combining variables $\mathbf{H}(x) + \mathbf{H}(y) \geq \mathbf{H}(x,y)$.

Table 4.1: Overview of basic statistical tests for one-sample and two-sample tests

test objective	single-sample	two-sample
bounding threshold	boundary tests, total received energy detector. Wisky's test	geometrical norms
parametric thresholds	Trend, Lillifors, Rao test	GLRT, Kullback, Wald
variable sample size	Wald sequential test	Cramer-Von Misses, Kolmogorov-Smirnov
amount of structure	histogram, self-information	Correlation, mutual entropy, mutual information

4.2.4 Dedicated filters

In many cases the hypothesis are formulated on the parameters θ that underlies certain distributions. There are two alternative strategies: one is to estimate θ from measurements and test by comparing, the other is a reference pattern associated with the hypothesized parameters that can be generated and compared to the data. There are two similar tests: the parametric correlator detector typically used in the time-domain, or the implementation using a FIR filter, where the filter coefficients are chosen to maximize sensitivity to the reference pattern. These approaches are generalized by the concept of classification.

Parametric correlator detector

Assuming the model $Y = \Phi^T\theta + E$, where $\text{cov}(E) = C$, and Y has a multi-variate normal distribution, a correlator detector is used to decide between the two hypotheses $H_0: \theta = 0$ vs. $H_1: \theta \neq 0$ considering an unknown parameter θ. The maximum likelihood estimator is then given by the Least Squares estimate, where $\bar{Y}$ is the pre-whitened data such that

$$\theta = (\Phi\Phi^T)^{-1}\Phi Y,\ \bar{Y} = \bar{\Phi}^T\theta + \bar{E} \text{ where } \bar{\Phi} = C^{-\frac{1}{2}}\Phi \text{ and } \text{cov}(\bar{E}) = I \tag{4.29}$$

The correlation detector [Wahlberg & Gustofsson, 2005] is given by

$$r = \sum_{i=1}^{N}\bar{y}(i)m(i) = \bar{Y}^T M \text{ where } M = \bar{\Phi}^T\theta \tag{4.30}$$

Matching filters

A matching filter is actually a two-sample test, where the known reference signal $m(t)$ of finite length N is convoluted with the new measurements to determine how much they resem-

ble. Thus the test statistic $r = l(N)$ based on the impulse response $H(i) = m(N-i)$ is used in the matching filter [Wahlberg & Gustofsson, 2005]:

$$l(t) = \sum_{0}^{t-1} h(i)\bar{y}(t-1) \tag{4.31}$$

Classifiers

A generalization of the correlation and matching filter detectors is the classifier. Given a number of different classes, representing the partition of the decision space into sections each associated with a hypothesis, a classifier maps measurement samples ξ to class memberships $\mu_c(\xi)$. The classifiers usually share projections, of the sample (input pattern). When the classes are limited to a signal of interest and unwanted signals (or noise), a optimal linear projection can be found to separate the two classes.

4.2.5 Projection methods

Static parity space approach

A parity function is an algebraic relationship involving observed measurements from the system such that the measurement noise is neglected [Frank, 1990]. The signal model is again $y = Cx + \Delta y$, where Δy is the $q \times 1$ error vector, y is the $n \times 1$ measured value vector, x is the actual value. For some threshold τ a fault is indicated in the ith measurement value if $\Delta y_i > \tau_i$. A set of parity equations is sought to obtain a $(q-n)$ dimensional parity vector p which can serve to compute a residual $r = V^T p$ [Frank, 1990]. Note that a redundant measurement is required to have $q > n$. These parity relations are in the simplest case linear independent: $p = Vy$. The design objective is to find the matrix V such that:

- The matrix V is a null space of C: $VC = 0$ and $V^T V = I_q - C(C^T C)^{-1} C^T$
- V is a unity matrix, i.e all it's vectors are orthogonal: $VV^T = I_{q-n}$

The residuals are then obtained by $r = y - C\hat{x}$ with $\hat{x}$ the least squares estimated given y, i.e. $\hat{x} = (C^T C)^{-1} Cy$. The residual is equal to $r = V^T p$. The transformation V applied to the signal model gives $Vy = VCx + V\Delta y$. Since $VC = 0$, the parity is $p = V\Delta y$, with $VV^T = I_{q-n}$ we will find:

$$r = V^T p = V^T V \Delta y = I_{q-n} \Delta y = \Delta y \tag{4.32}$$

Subspace projection nulling

This method is applied in radar processing and radio frequency interference (RFI) mitigation [Boonstra, 2005]. Assuming an array vector $\boldsymbol{x}(t)$ to consist of a signal of interest (or a nuisance strong interfering signal) and noise: $\boldsymbol{x}(t) = A\boldsymbol{s}(t) + \boldsymbol{n}(t)$. The cross-correlation of the array vectors is $R(t) = x(t)x^T(t)$. The gain matrix A relates the signals to the sensors.

$$R = E\{x(t)x^H(t)\} = AE\{\boldsymbol{s}(t)\boldsymbol{s}^H(t)\}A^H + E\{\boldsymbol{n}(t)\boldsymbol{n}^H(t)\} \tag{4.33}$$

We define $P = E\{\boldsymbol{s}(t)\boldsymbol{s}^H(t)\}$. Now the goal is to find the range space of the signals $\boldsymbol{s}(t)$. This is achieved by separating the signal from the noise space with Eigen decomposition:

$$R(t) = APA^H + \sigma^2 I = U\Lambda U^T = U_s\Lambda_s U_s^T + U_n\Lambda_n U_n^T \qquad (4.34)$$

Using σ^2 as a threshold, the Eigenvectors can be split into signals for which $\lambda_i \geq \sigma^2$ and noise for which $\lambda_i < \sigma^2$. Then the corresponding Eigenvectors are kept in either U_s or in U_n for signal and noise resp. Now the non-zero columns of U_s span the range space of A. To cancel the signals from the observed array vectors the objectives is to project out the signals through a transformation $\boldsymbol{x}_n(t) = P_v^{\perp}\boldsymbol{x}(t)$. The method for doing this is by concatenating the Eigenvectors corresponding to the signals to cancel (i.e. those with $\lambda_i \geq \sigma^2$) columns in a matrix $U_s = (\boldsymbol{u}_{s1}, \ldots \boldsymbol{u}_{sk})$, assuming k Eigenvectors correspond to signals to cancel. Then the projection matrix is found through a least squares solution to $P_v^{\perp}\boldsymbol{u}_{si} = \boldsymbol{0}$

$$P_v^{\perp} = I - U_s(U_s{}^H U_s)^{-1} U_s{}^H \qquad (4.35)$$

Principal component analysis

Principal Components Analysis (PCA) operates directly on a block matrix of sequential multivariate measurements. A projection matrix V can be derived such that the columns provide the orthogonal basis onto which the data is projected. The axes in the new space are ordered by increasing variance. These variances are actually the singular values and the basis vectors or the singular vectors of a Singular Value Decomposition discussed in appendix B.1. Similar to the parity space approach a linear relationship between the variables is used to separate the noise and signal space. However the difference is that with the parity space approach the expected relationship is known to be $Y = CX$. In case of PCA this is neither known nor assumed, hence it is a blind projection method.

4.2.6 Adaptive Filters

The dynamic linear relationships are be expressed by moving average (MA) and auto-regressive (AR) parameters. The on-line estimation of the ARMA parameters from observations is a key technique for detection of change in the dynamic relationships. These estimations are implemented by adaptive filters.

Least squares fit

Objective is to determine a and b from the measurements of y and u considering the system $y(t) = ay(t-1) + bu(t-1) + e(t)$. The cost-function is $J(a,b) = \sum (y(t) + ay(t-1) - bu(t-1))^2$. It follows that the least-squares solution is:

$$\begin{bmatrix}\hat{a}\\ \hat{b}\end{bmatrix} = (\Phi^T\Phi)^{-1}\Phi^T\begin{bmatrix}y(2)\\ y(3)\\ \ldots\\ y(4)\end{bmatrix}, \text{ where } \Phi = \begin{bmatrix}y(1) & u(1)\\ y(2) & u(2)\\ \ldots & \ldots\\ y(N-1) & u(N-1)\end{bmatrix} \qquad (4.36)$$

More elaborate filters exist for the estimation of ARMA parameters, also in iterative forms, (for instance the iterative recursive least squares method listed in chapter 3). Particularly the stable estimation of the AR parameters is difficult. Auto-correlation and the least-squares approach is the most straightforward approach. Alternatives use a spectral analysis or maximum entropy. Well-known implementations are Yuler-Walker, Levison-Durbin and Burg's algorithm [Brockwell and Davis, 1986].

4.2.7 Design, quality and optimality

Some basic design and optimization procedures from signal detection theory conclude this overview on statistical signal detection. The key issue in most of the above statistics is how to choose the decision threshold. Theoretical procedures to determine the optimal threshold are the Neyman-Pearson strategy and the Bayes detector. Complementary to these methods some graphical methods are available to optimize if parametric distributions are known a priori or estimated from examples. Optimization and verification requires a comparison among different tests. Some regularity conditions have been proposed by Capon to allow for such a comparison (appendix D.4).

Theoretical optimal threshold.

The Neyman-Pearson strategy [Hancock and Wintz, 1966; Cheng, 2004; Wahlberg and Gustofsson, 2005] is used for a fixed number of samples. It finds a optimal threshold, by fixing either α or β. The quality of statistical tests is often expressed with the power $1-\beta$ which is the detection probability of accepting normality when the information source is in fact normal. The literature on traditional and parametric approaches tends to take H_0 as the "normal" situation, i.e. the probability of false acceptance β, $P(H0|H_1)$ is usually fixed.

$$L(\boldsymbol{x}) = \frac{f(\boldsymbol{x}|K)}{f(\boldsymbol{x}|H)} \begin{matrix} > \\ < \end{matrix} \lambda \Rightarrow \begin{matrix} \text{accept } K \\ \text{accept } H \end{matrix} \tag{4.37}$$

Bayes detection is pursuing a cost-based risk minimization. Each of the cases for detection is associated to cost (C_{HH} is the cost of accepting H when H is true, while C_{KH} is the cost of accepting K when H is true, etc.). The Bayes detector is also written as

$$P(H)(C_{KH}-C_{HH})f(x|H) \begin{matrix} > \\ < \end{matrix} P(K)(C_{HK}-C_{KK})f(x|K) \Rightarrow \begin{matrix} \text{accept } K \\ \text{accept } H \end{matrix} \tag{4.38}$$

Assuming $C_{HK}-C_{KK}>0$, this is written as the likelihood ratio $L(\boldsymbol{x})$ and threshold of the test τ_B:

$$L(\boldsymbol{x}) = \frac{f(\boldsymbol{x}|K)}{f(\boldsymbol{x}|H)} \begin{matrix} > \\ < \end{matrix} \frac{P(H)(C_{KH}-C_{HH})f(x|H)}{P(K)(C_{HK}-C_{KK})f(x|K)} = \tau_B \tag{4.39}$$

Graphical methods

A plot of the detection probability $1-\beta$ versus the false alarm probability α is sometimes useful in describing the performance of a detector. Plots of this type are called *Receiver Operating Characteristics* (ROC). Given a detector on a random variable $D(\boldsymbol{x})$ taking values 0 and 1, H is accepted if $D(\boldsymbol{x}) = 0$ and else K. The operating characteristic of $D(\boldsymbol{x})$ only depends on the distribution $F \in K$ of $\boldsymbol{x}$. Hence ROC for a detector $D(\boldsymbol{x})$ is denoted $Q_D(F) = \mathrm{E}(D(x))$. The plot is obtained from $Q_D(F)$ by varying distributions of $\boldsymbol{x}$ obtaining $\alpha = Q_D(F \in H)$ and $1-\beta = Q_D(F \in K)$. All clever tests give a curve above this straight line. Alternative plots are

- Detection performance: $1-\beta$ versus SNR for a given α. Again this performance is maximized for the Neyman-Pearson test.
- BER: Bit-Error Rate. In communication, both hypotheses are equal, and the design is to get $1-\beta$ versus α. A BER-plot shows the $1-\beta = \alpha$ point for different SNR.

Verification and selection

In non-parametric design of detectors some regularity conditions [Gibson and Melsea, 1991] have been derived by Capon to verify asymptotic optimality for the detection of signals in noise; we have included these in appendix D.4. These regularities on statistics ensure the possibility to compare statistical tests, selecting in the design the best among a class of detectors. A way to compare tests is the number of samples required to decide between the H_0 hypothesis and any alternative, when they are increasingly resembling each other e.g. asymptotic relative efficiency (ARE).

4.3 Fault detection and isolation

4.3.1 Preliminaries

The statistical signal detection techniques are suitable for communication and sensing. However, they ignore the structure and state of the underlying information source. Since we are considering controlled systems we can significantly improve the system behavior model by describing the dynamics of the data generating process, using knowledge of the physical system, the so-called first principles. The exact model of the essential process, excluding the enforced control or observers derived from the first principles is called the ***nominal process model***. Assume the process is a controlled system, a controller is required to be designed or modelled, and a state observer measuring the knowledge of the exact state is required to derive the appropriate actions. The system-oriented and model-based domain of FDI takes on the challenge of modeling the states and state transitions in the information source. The modeling of controlled systems in FDI relies on control systems theory introduced in section 2.4.

4.3.2 Dedicated filters

The basic principle in FDI is to compare presumed behavior of the system with actual behavior through known properties of the data generating process. Algebraic manipulation of the system equations are combined with statistical techniques to design an optimal tests for the detection of disturbances. Similar to the dedicated filters in signal detection, there are two approaches: 1) the comparison of properties which are estimated from data (in FDI a model is used to estimate the properties); 2) the design of a dedicated fault filter or observer optimized for specific reference models of faults and disturbances.

Parameter identification approach

The reference approach in FDI is the framework defined by Isermann [Isermann, 1984; Frank, 1990]. This approach is to identify changes in non-measurable physical quantities (NMQ) though the knowledge of the physical principle. Fault detection, based on process coefficients and features which are mostly not directly measurable quantities (NMQ) require on-line parameter estimation methodsI [Isermann, 1984]. The essential steps are:

1. Choose a parametric model of the systems, in the normal case:

$$a_n y^{(n)}(t) + \ldots + a_n \dot{y}(t) + y(t) = b_0 u(y) + \ldots + b_m u^{(m)}(t) \qquad (4.40)$$

2. Determine the relationships between the properties p_j and the model parameters θ_i

$$\theta = f(p) \qquad (4.41)$$

3. Identify the model parameter θ from the input u and output y of the actual system
4. Determine the physical parameter vector

$$p = f^{\text{inv}}(\theta) \tag{4.42}$$

5. Calculate the vector of deviations Δp from the values taken from the nominal model.
6. Estimate the fault from the relationship between faults and physical changes Δp_i.

Dedicated observers

A closed-loop parity space approach leads naturally to state estimation [Frank, 1990]. The motivation of a state space approximation approach is: 1) to compensate for differences in initial conditions; 2) to stabilize the model in the presence of an unstable system; 3) to provide freedom in design of the detection filters so that the fault effects can be decoupled. Dedicated observers aim to estimate the state of a system to find drift in the state space. Hence the model is assumed to remain applicable despite linearization in an equilibrium no longer being enforced. Dedicated observers are actually a form of adaptive filter, discussed below.

Fault detection filters

We have discussed several fault models in section 4.1.2. The optimal observers for known faults can be derived through some algebraic manipulation. Assume the model in equation 4.43 for a known system and known fault profiles [Wilsky, 1976; Aires, 1999], while k_i is the fault direction for $i = 1 \ldots r$ different fault directions and $f_i(t)$ is an arbitrary scalar function. Then k_i corresponds to actuator faults, and k_j to sensor faults.

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + k_i f_i(t) \\ y(t) &= Cx(t) + \bar{k}_j \bar{f}_j(t) \end{aligned} \tag{4.43}$$

For the optimal H (discussed in chapter 2), the state observer equations becomes

$$\begin{aligned} \dot{\hat{x}} &= (A - HC)\hat{x} + Bu + Hy \\ \hat{y} &= C\hat{x} \end{aligned} \tag{4.44}$$

Actuator faults (a) and sensor faults (b) are associated with the state error according to

$$\text{a)} \quad \begin{aligned} \dot{\varepsilon} &= (A - HX)\varepsilon + k_i f_i \\ r &= C\varepsilon \end{aligned} \qquad \text{b)} \quad \begin{aligned} \dot{\varepsilon} &= (A - HX)\varepsilon + k_j f_j \\ r &= C\varepsilon_j + \bar{k}_j \bar{f}_j \end{aligned} \tag{4.45}$$

4.3.3 Projection methods

There are state-space projection methods similar to the static parity space approach and subspace projection nulling methods. The projection, identifying deviations between actual input-output trajectories and the nominal model, is given by the dynamic parity space approach [Frank, 1990]. Given a fault model which includes sensor and actuator faults, a similar robust time-domain approximation can be derived. This method is called robust as the different types of disturbances (state, sensor and actuator deviations) are decoupled.

Dynamic parity space approach

Assuming the nominal linear system $x(k+1) = Ax(k) + Bu(k)$ and $y(k+1) = Cx(k)$, with x the $n \times 1$ state vector, and u the $p \times 1$ actuator input, and y the $q \times 1$ sensor output vector. The redundancy relations can be specified. First of all the $(s+1)q$ dimensional vectors v are given by

$$P = [v | v^{\mathrm{T}}\Gamma = 0] \qquad \text{with} \begin{bmatrix} C \\ CA \\ \cdots \\ CA^s \end{bmatrix} = \Gamma_s \tag{4.46}$$

This is called the parity space of order s. Every parity vector v can be used at any time k for a parity check, generating a residual $\hat{r}(k)$. H is the Hankel matrix as specified in appendix D.4.

$$r(k) = v^T \left[\begin{bmatrix} y(k-s) \\ \cdots \\ y(k) \end{bmatrix} - H \begin{bmatrix} u(k-s) \\ \cdots \\ u(k) \end{bmatrix} \right] \tag{4.47}$$

The residuals are then obtained by filling the state equations, with Γ as observability matrix:

$$r[k] = v^{\mathrm{T}}\Gamma \boldsymbol{x}[k-s] \tag{4.48}$$

A robust time-domain approximation

The optimal time-domain approximation [Frank, 1990] is found by unfolding the state-space equations up to a finite time-horizon s. Then the equations 4.49 and 4.50 are found.

$$\begin{bmatrix} y_{k-s} \\ y_{k-s+1} \\ \cdots \\ y_k \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \cdots \\ CA^s \end{bmatrix} x_{k-s} + H_1 \begin{bmatrix} u_{k-s} \\ \boldsymbol{u_{k-s+1}} \\ \cdots \\ u_k \end{bmatrix} + H_2 \begin{bmatrix} d_{k-s} \\ \boldsymbol{d_{k-s+1}} \\ \cdots \\ d_k \end{bmatrix} + H_3 \begin{bmatrix} f_{k-s} \\ \boldsymbol{f_{k-s+1}} \\ \cdots \\ f_k \end{bmatrix} \tag{4.49}$$

The Hankel matrices H_1, H_2 and H_3 are given by

$$H_1 = \begin{bmatrix} 0 & & & \\ CB & 0 & & 0 \\ CAB & CB & 0 & \\ \cdots & \cdots & \cdots & 0 \\ CA^{s-1}B & \cdots & \cdots & CB\ 0 \end{bmatrix} \quad H_2 = \begin{bmatrix} F & & & \\ CE & F & & 0 \\ CAE & CE & F & \\ \cdots & \cdots & \cdots & F \\ CA^{s-1}E & \cdots & \cdots & CE\ F \end{bmatrix} \quad H_3 = \begin{bmatrix} G & & & \\ CK & G & & 0 \\ CAK & CK & G & \\ \cdots & \cdots & \cdots & G \\ CA^{s-1}K & \cdots & \cdots & CK\ G \end{bmatrix} \tag{4.50}$$

The scalar residual to be generated has to check whether the above state equations hold for the available input and output data. This is done by calculating on-line for each sample time k:

$$r_k = v^T \left[\begin{bmatrix} y_{k-s} \\ y_{k-s+1} \\ \cdots \\ y_k \end{bmatrix} - H_1 \begin{bmatrix} u_{k-s} \\ \boldsymbol{u_{k-s+1}} \\ \cdots \\ u_k \end{bmatrix} \right] \text{ where } \forall v : v^T \begin{bmatrix} C \\ CA \\ \cdots \\ CA^s \end{bmatrix} = 0 \qquad (4.51)$$

Since the residual has been affected by the fault, v must be determined to meet $v^T H_3 \neq 0$. Also v has to meet $v^T H_2 = 0$ as the residual should not be affected by an unknown input vector d. The performance index to optimize can be chosen as $P = \|v^T H_2\| / \|v^T H_3\|$. Now if the matrix V_0 is the basis for the space of all solutions v, the design problem can be formulated as to find the vector w which minimizes the performance index

$$P = \frac{\|w^T V_0 H_2\|}{\|w^T V_0 H_3\|} \qquad (4.52)$$

The solution can be found by differentiations w.r.t. w leading to equation 4.53. This is a general Eigenvector/Eigenvalue problem which is solved by taking for the optimal residual vector $v = w$ the Eigenvector corresponding to the smallest Eigenvalue.

$$w^T (V_0 H_2 H_2^T V_0^T - P V_0 H_3 H_3^T V_0^T) = 0 \qquad (4.53)$$

4.3.4 State estimation through adaptive filtering

In the state-space model, the state vector determines the dynamic relation between input and output. The estimation of the state from the input-output data is a type of adaptive filtering, be it that it is restricted to the mapping defined by the system matrices A, B, C and D. We discuss three adaptive filtering solutions, each fit for certain faults and systems, see table 4.2.

Table 4.2: Adaptive filter approaches

System State Transitions	Dependencies Between Fault and State	Proposed Solution
Linear	Linear	Kalman Filter
Linear	Non-Linear	Non-linear Decoupling
Non-linear	Non-Linear	Extended Kalman Filter

Kalman filter

The goal is to estimate the state $x \in \Re^n$ of a process with a linear stochastic difference equation (equation 4.54). with both w_k and v_k i.i.d zero-mean normally distributed, i.e. $p(w) = N(0, Q)$ and $p(v) = N(0, R)$. One defines a priori state estimates and a posteriori state estimates, with $\hat{x}_k^-$ the a priori state and $\hat{x}_k$ the a posteriori state. The Kalman Filter (appendix C) consists of: 1) a prediction step, and 2) measurement feedback update [Welch & Bischop, 2004], these steps are shown in table 4.3.

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \text{ and } z_k = Cx_k + v_k \qquad (4.54)$$

Table 4.3: The prediction and feedback in the Kalman filter.

Prediction Step	Feedback Step
$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}$ $P_k^- = AP_{k-1}A^T + Q$	$K_k = P_k^- CX^T(CP_k^- C^T - R)^{-1}$ $\hat{x}_k = \hat{x}_k^- + K(z_k - C\hat{x}_k^-)$ $P_k = (I - K_k C)P_k^-$

Non-linear decoupling

A difference in the earlier state system models is found in the linearity of the input-output relations, while all system models assume a linear state transition. In non-linear decoupling a transformation T is sought to optimize an error response to the non-linear input-output relations, i.e. assuming the model described in equation 4.55.

$$\begin{aligned} \dot{x} &= Ax + B(y,u) + Ed + K(x)f \\ y &= Cx + G(x)f \end{aligned} \tag{4.55}$$

A corresponding fault observer is then given by

$$\begin{aligned} \dot{z} &= Rz + J(y,u) + Sy \\ r &= L_1 z + L_2 y \end{aligned} \tag{4.56}$$

Assuming the absence of faults z results from a linear transformation of x: $z = Tx$. Then the estimation error (equation 4.57) has to be solved for T, thus solving the set of equations 4.58.

$$\begin{aligned} \dot{e} &= \dot{z} - T\dot{x} \\ &= Rz + J(y,u) + Sy - TAx - TB(y,u) - TEd - TK(x)f \end{aligned} \tag{4.57}$$

$$\begin{aligned} TA - RT &= SC \\ TE &= 0 \\ J(y,u) &= TB(y,u) \\ L_1 T + L_2 C &= 0 \end{aligned} \tag{4.58}$$

Extended Kalman filtering

A solution to the non-linear input-output model with non-linear state transition is the Extended Kalman Filter [Welch & Bischop, 2004]. It is based on a linearization of the system in the equilibrium, assuming a system model:

$$x_k = f(x_{k-1}, u_{k-1}, w_{k-1'}) \text{ and } z_k = h(x_k, v_k) \tag{4.59}$$

In appendix C.3 we derive the result shown in table 4.4 [Welch & Bischop, 2004]. The EKF solvability depends on a reasonable estimate of the initial error covariance estimate P_0 and blending matrix K_0, and also on the observability of the system.

Table 4.4: The prediction and feedback in the Extended Kalman Filter

Prediction Step	Feedback Step
$\hat{x}_k^- = f(\hat{x}_{k-1}, u_{k-1}, 0)$ $P_k^- = A_k P_{k-1} A_k^T + W_k Q_{k-1} W_k^T$	$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + V_k R_k V_k^T)^{-1}$ $\hat{x}_k = \hat{x}_k^- + K_k(z_k - h(\hat{x}_k^-, 0))$ $P_k = (I - K_k H_k) P_k^-$

Process monitoring

Assuming the general state-space model, with linear state transition, the goal is to learn/model the fault functions f_x and f_y [Trunov and Polycarpou, 2000]. The tracking of fault functions offers an elegant way to distinguish between different rates of change, which can be regulated by a learning parameter.

4.3.5 Blind Identification

The need for blind identification of the system matrices arises in cases where one assumes a system with state, possibly starting from an initial state-space model, while the system is not considered time-invariant. The objectives of blind identification of the linear system is to estimate the system matrices A, B, C and D, assuming the typical linear model $x(k+1) = Ax(k) + Bu(k)$ and $y(k) = Cx(k) + Du(k)$. The following solutions are described in appendix D.5.

Markov parameters

The *Hankel matrices* of Markov parameters are defined as

$$H(k) = \begin{bmatrix} \Upsilon_k & \Upsilon_{k+1} & \cdots & \Upsilon_{k+\beta-1} \\ \Upsilon_{k+1} & \Upsilon_{k+2} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \Upsilon_{k+\alpha-1} & \cdots & \cdots & \Upsilon_{k+\alpha+\beta-2} \end{bmatrix} \tag{4.60}$$

where, Υ_0, Υ_1,... are *Markov parameters* given by $\Upsilon_0 = D$ $\Upsilon_1 = CB$, $\Upsilon_i = CA^{i-1}B$, for $i = 2, 3, \ldots$. The Markov parameters Υ_0, Υ_1, are constructed from given impulse responses without explicit knowledge of the system matrices A, B, C, and D. If the order of the system to be identified is n, then the choices of $\alpha \geq n$ and $\beta \geq n$ ensure that the matrix is of rank n.

If the singular value decomposition of the Hankel matrix is $H(0) = U\Sigma V^T$, with Σ a diagonal matrix U and V both unitary ($UU^T = I$), then the matrices of the minimal state-space realization can be estimated with:

$$A = \Sigma^{-\frac{1}{2}} U^T H(1) V \Sigma^{-\frac{1}{2}} \tag{4.61}$$

$$B = \Sigma^{\frac{1}{2}} V^T \begin{bmatrix} I_p \\ 0 \\ \dots \\ 0 \end{bmatrix} \tag{4.62}$$

$$C = \begin{bmatrix} I_p & 0 & \dots & 0 \end{bmatrix} U \Sigma^{\frac{1}{2}} \text{ and } D = \Upsilon_0 \tag{4.63}$$

State sequence generation through trajectory fitting

This state sequence generation through trajectory fitting corresponds to the Markov parameter approach above. Now two new block Hankel matrices are defined using only I/O measurements [Moonen M., De Moor B., Vandenberghe L. and Vandewalle J.,1989]:

$$H_1 = \begin{bmatrix} Y_{h1} \\ U_{h2} \end{bmatrix} \text{ and } H_2 = \begin{bmatrix} Y_{h2} \\ U_{h2} \end{bmatrix} \tag{4.64}$$

where

$$Y_{h1} = \begin{bmatrix} y[k] & \dots & y[k+j-1] \\ \dots & \dots & \dots \\ y[k+i-1] & \dots & y[k+j+i-2] \end{bmatrix} \tag{4.65}$$

$$Y_{h2} = \begin{bmatrix} y[k+i] & \dots & y[k+i+j-1] \\ \dots & \dots & \dots \\ y[k+2i-1] & \dots & y[k+j+2i-2] \end{bmatrix} \tag{4.66}$$

U_{h1} and U_{h2} are constructed similarly. The main theory applied in this approach is the relationship between these matrices and the state vectors:

$$Y_{h1} = \Gamma_i . X_1 + H_t . U_{h1} \tag{4.67}$$

$$Y_{h2} = \Gamma_i . X_2 + H_t . U_{h2} \tag{4.68}$$

where $X_2 = \begin{bmatrix} x[k+i] & \dots & x[k+i+j-1] \end{bmatrix}$. The state sequence generation for X_2 is then estimated, using the pseudo-inverse Γ_i^+ such that $\Gamma_i^+ \Gamma_i = I$, by:

$$X_2 = \Gamma_i^+ . Y_{h2} - \Gamma_i^+ . H_t . U_{h2} = \begin{bmatrix} \Gamma_i^+ & -\Gamma_i^+ . H_t \end{bmatrix} . \begin{bmatrix} Y_{h2} \\ U_{h2} \end{bmatrix} \tag{4.69}$$

A computational reduction is achieved by the SVD of the Hankel matrices, see appendix D.5.

4.3.6 Selecting an FDI strategy

So far we have presented the basic concepts in the detection of disturbances and faults using a dynamic model of the system behavior, which is the parameter identification approach formulated by Isermann. We have organized the different solution by the required knowledge of the faults and the systems with their interaction

- Known faults with a known system model can be detected through static filters, using only the output response of dedicated fault detection filters, or through classifiers based on the output error of the system or through dedicated filters.
- Parameterized faults and systems can be addressed by dedicated observers, but require parameter estimation. If the system and the disturbances are presumed to be independent and one has information on the statistical properties and dynamic of the faults, the statistical signal detection techniques discussed in the previous section can be applied.
- When the system and disturbances interact but the relationships are known, dedicated observers are required to estimate and update the state estimate. In section 4.3.4 we discuss Kalman filters, EKF, and non-linear decoupling.
- In case the system behavior during normal operation is known, but not much specific is known about the disturbances and faults one can use projection methods. Specifically we will describe the dynamic parity space approach, which is the dynamic analogon for the static parity space approach discussed in the previous section. A robust time-domain extension for actuator and sensor disturbances is presented, defining Hankel matrices not only for the nominal state space matrices but also for the sensor and actuator faults according to the system fault model.

In case the interaction between faults and system is such that the system model is no longer invariant, a system identification approach can be used to estimate and update the system model. We discussed the Markov estimate, and a direct fit of input-output behavior. This is the state-space analogon of fitting an ARMA model.

4.4 Computational intelligence

4.4.1 Preliminaries

Computational Intelligence (CI) is the discipline which deals with algorithms mimicing intelligent behavior in humans and biology, specifically algorithms in CI mimic adaptation and inference. There are two interesting intersections between computational intelligence and detection: methods applied to search and diagnose in a more intelligent way and second alternative models and projections using neural networks. Computational intelligence, being inspired from natural behavior, has a less firm theoretical foundation compared to statistics and system theory. They offer ways to deal with challenges that are not tackled by the conventional disciplines. We identify some particular challenges:

- **Complexity**. In the case of qualitative model-based approaches, the combinatorial complexity is unavoidable and can only be partly alleviated with efficient search [de Kleer & Brown, 1984; Reiter, 1987]. Because of the combinatorics many multiple fault combinations arise and the search for multiple faults by specifying them explicitly as different classes and obtaining training patterns is not feasible [Venkatasubramanian, 2003].
- **Open rather than closed search space**. Correlation detection of whitened signals with stored replica's of the signal alphabet, is an optimal strategy for detecting deterministic signals in additive noise [Hancock & Wintz, 1966]. However expressing system requirements involves specifying against unwanted behavior in response to unforeseen events while many applications are now targeting environments that cannot be considered as closed and for which knowledge representations will necessarily be incomplete [Lisboa,

2000]. Since there are only partial specifications of possible behaviors the search space becomes an open space and unbounded space.

- **Non-smooth parameter spaces**. Often a modeling or detection problem is only partially specified: an exact model is not available and the state space is not completely known. Consequently singularities occur and one faces a limited observability. Linear algebra, statistical solvers and iterative parameter fitting methods fail.
- **Non-cardinal values**. In many real-world situations there is information which is not numerical in nature, e.g. when there are different materials playing a part in a process, these materials are affecting the behavior but they are not numerically represented. A symbolic representation is possible. However, it is hard to incorporate knowledge related to these non-cardinal variables in a quantitative approach.
- **Expert knowledge**. In general human expert knowledge is often available though hard to capture with a statistical or mathematical model. Qualitative information and implicit knowledge however can be the key to detection. Particularly human reasoning is very effective in diagnosis. Moreover humans have trouble in understanding the complex mathematical representations of behavior; rules are much easier to understand.

4.4.2 Search and diagnostic methods

Relating symptoms to effects

A good discussion on search and diagnostic methods can be found in [Venkatasubramanian, 2004]. Diagnostis is based on transforming symptoms to fault estimates. These transformations use (a) direct lookup, (b) causal model-based reasoning (deep knowledge) or (c) process history (shallow knowledge). There are three transformation steps [Venkatasubramanian, 2003]

1. measurement space to feature space;
2. feature space to decision space;
3. decision space to [fault] class space;

Simulation and diagnosis are two generic ways to compare measured behavior to expected structure, which is available either as a nominal process model or as fault patterns. Diagnosis is finding the structure in behavior [Venkatasubramanian, 2003], i.e. find the relationships between disturbances and faults out of the residual. Simulation is finding behavior from structure [Venkatasubramanian, 2003], given a model of the system and/or the faults generate examples of the possible behavior. An example of simulation is perturbation analysis. In perturbation analysis, examples of abnormalities are generated by perturbing the parameters of a system model. The causal relationships between symptoms, failures and faults can be a very complex. They are not easily described by exact rules. Yet during system operation one gathers vast amounts of examples, e.g. execution traces. Computational intelligence offers black-box learning methods to fit classifiers from measurements labelled by associated failures, i.e to classify faults by their symptoms [Rault and Baskiotis, 1989]. There is a distinction between relationships that can be known without actually knowing the system and relationships that are known from the physical or logical system principles. This distinction is expressed as shallow knowledge versus deep knowledge. The latter corresponds to the actual system, i.e. understanding by a white-box model.

Closed versus open search

There are two different types of search methods for effective hypothesis generation:

- **topographic search** performs malfunction analysis using a template of normal operation from a composite system model.
- **symptomatic search** looks for symptoms to direct the search to the fault location [Venkatasubramanian, 2003], through look-up tables or hypothesize-and-test-search.

In symptomatic search there is a key difference between open-loop and closed-loop approaches for hypothesis generation. In closed-loop search there is a single adaptive reference model, implying an open/unbounded set of models. Such a search is non-deterministic in nature, i.e. the behaviors are not determined a priori and new hypothesis' are generated during the search. It is also called a non-decision directed measurement [Hancock & Wintz, 1966] in the context of signal detection, i.e. the signal to detect is to be identified rather than known a priori. In an open-loop search there is/are non-adaptive model(s), i.e the hypothesis is defined by a finite set of reference models, making the search deterministic. This is known as decision directed measurement [Hancock & Wintz, 1966], as the specific signals to be received/ detected are a priori known.

Intelligent search methods

Intelligent search methods are those beyond the capabilities of quantitative/numerical: hypothesis tests, such as the discussed test statistics: Wald, GLRT, Kullback, Generalized Jensen Difference Measure. Improvements are achieved by incorporating qualitative information, (human expert knowledge and experiences from case studies) or through translation of numerical results into the human expert reasoning domain. Particularly, symbolic representation helps to handle non-cardinal variables which cannot be effectively used in quantitative approaches. The pure qualitative methods are expert systems, fault dictionary, and diagnostic trees. However neglecting numerical observation greatly reduces the accuracy of an approach and the optimality of a performance as numerical verification is not possible. Hybrid methods combine qualitative and quantitative models. Examples of hybrid methods are fuzzy logic and belief networks [Horvitz et al., 1988].

4.4.3 Applications of neural networks in detection

The role of neural networks with respect to detection, isolation and diagnosis cannot be ignored in the context of this thesis, but we like to emphasize again that using neural networks for detection is not the research subject in this thesis. For more information, one is referred to [Veelen, 2000a]. The neural model can be a process model or a fault model. Though there is a wide variety of neural networks appearing in detection problems, successful applications are limited to a few strategies. In principle the neural model can be used (like any other type of model) to detect, monitor, isolate, diagnose and accommodate. However the use of neural networks in detection is almost always in combination with a more conventional exact or statistical modeling approach. Neural networks are considered useful complementary models, since according to [Lisboa, 2001]:

> *"expressing system requirements involves specifying against unwanted behavior in responses to unforeseen events while many applications are now targeting environments that cannot be considered as closed and for which knowledge representations will necessarily be incomplete."*

There are four application strategies for neural networks in detection: 1) behavioral modeling of normal behavior through clustering; 2) process monitoring through clustering; 3) tracking; and, 4) novelty detection, see figure 4.3.

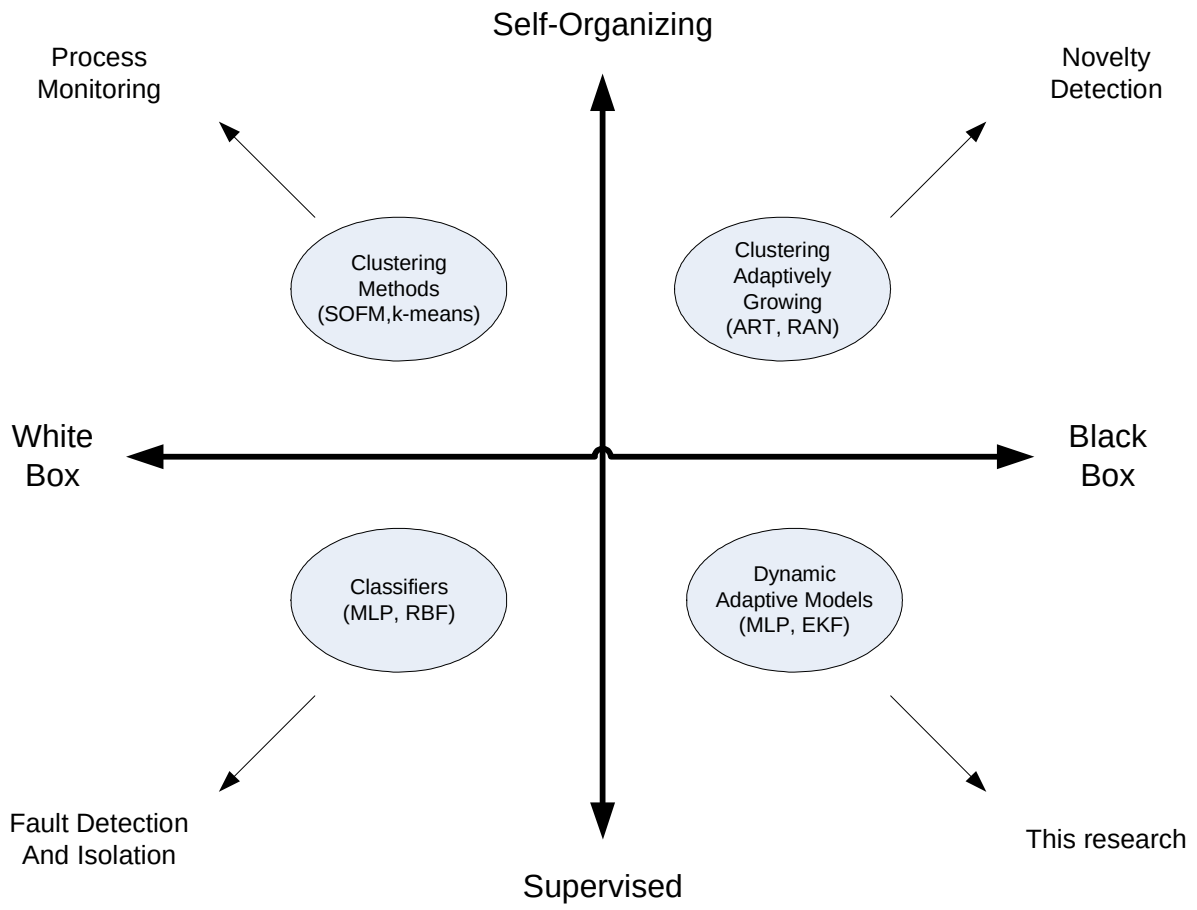

Figure 4.3 : Different applications of neural networks for detection

The modeling of normal behavior directly from data through clustering. In this case faults and disturbances are measured by the distance between data observed from estimated distributions, and data measured under normal operating conditions. The distributions are expressed by kernels of a clustering neural network such as the Kohonen map or ART [Yen and Feng, 2001], [Wong, 1997], [Desforges et al., 1998]. Clustering methods are particularly useful for sensor fault detection when dealing with many possibly heterogeneous sensor signals. They reduce the complexity through data fusion [Taylor, 2000], and by identifying essential patterns in data i.e. data mining [Fayyad, 1996].

The monitoring for detection of process parameters or process mode/state variables. Typically in an industrial process there are some unmeasurable process coefficients that cannot be directly estimated, but which can be derived from the data or from state variables that are tracked by a classical system theoretical observer. Monitoring tests whether coefficients are within certain boundaries to guarantee proper operation. The boundaries are either estimated by clustering or monitored by tracking coefficients [Tseng and Chou, 2002].

The tracking and corrective adaptation of smooth (incipient) changes. A correction factor, typically additive and/or multiplicative, is used to correct a nominal process model. The factor is estimated from data with inputs, that are set-points, conditions, and the error made by the nominal process model [Rengaswamy, 2000]. The estimations made by the neural networks can be improved through on-line learning. Adaptive and learning methods have already been suggested for an industrial setting by [Hancock and Wintz, 1966].

The detection of additional independent abnormal signal components. Signal components that do not interfere with known behavior (dependencies that are already known) are detected by two methods. The first method is to use a known fault model matched against observed data [Hummels, 1995]. The second method is to detect and isolate new signal components. To this purpose one can use neural networks in the fashion of blind source separation, independent component analysis or auto-association. Non-linear projection methods help to optimize the signature computation, since most detection problems are not easily reduced to a linear projection problem. We can also find solutions on the area of project pursuit [Mao & Jain, 1995].

Neural networks are used as a classifier of known faults or (in case of novelty detection) for statistical residue-based detection. A third neural application is the clustering of acceptable process parameters estimated through a physical-principle model. This shows that neural networks are applied in conventional ways, i.e. the neural network does not replace conventional nominal process models and disturbances as if they are independent of the system.

4.5 Discussion

In this chapter we have provided an overview of the state-of-the-art techniques contributed from different disciplines. Some of these techniques are essential for detection. However they may require an extension to deal with new applications and new types of systems. We can now explain the problem domain of this research, by categorizing the available techniques according to the modeling approach required for the system and it’s abnormalities. In section 4.5.1 we abstract from the techniques specific to each discipline to the more general mechanism behind the techniques. In section 4.5.2 we organize the mechanisms and indicate where problems arise. These problems are investigated in the next chapter.

4.5.1 Overview of the techniques organized by underlying mechanisms

In section 4.1.5 we have discussed the key functions and basic techniques from statistics, linear algebra and control systems theory. We have explored beyond the base technology in the disciplines of signal detection, FDI and computational intelligence. This literature survey covers a wide spectrum of techniques from these disciplines. The depth of this survey is limited to the illustration of some key mechanisms provided in each domain. The different mechanisms which appear to be similar in FDI and signal detection are: dedicated filtering, projection methods, adaptive filtering and blind identification. Figure 4.4 gives an overview of techniques categorized by the mechanisms.

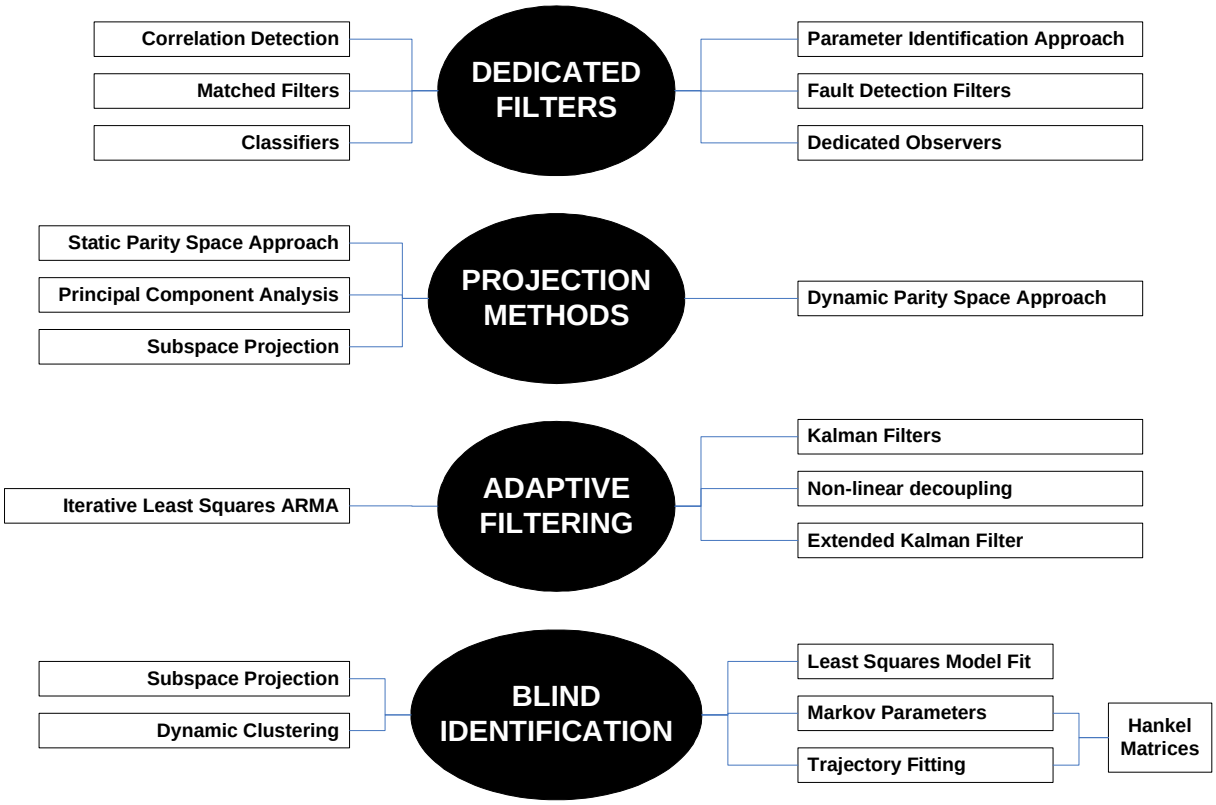

Figure 4.4 : Overview of detection techniques: residual based (left) and parameter based (right).

The design challenge in detection is to apply the right mechanisms with the right technique given the problem at hand. The properties of the application, system and abnormalities, lead to a detection strategy to follow. We have discussed the choice of technique per discipline. Table 4.1 specifies when to use which basic statistical test. This can be used for test design after signature computation. In subsection 4.3.6 we have discussed the selection of an appropriate FDI strategy given the knowledge of a system and it's faults. Similarly subsection 4.4.3 describes how neural networks can be applied for detection. We now take a more abstract view regarding the adequacy of the mechanisms in relation to the systems and applications.

4.5.2 Problem domain

The toughest detection issues arise in the modeling and estimation of system and faults, particularly due to dependencies and variability. The existence of unknown or hidden cross-dependencies between the process signals is one of the major causes of upsets and accidents in the process industry today [EEMUA, 1999]. The need for detection comes from variability [Venkatasubramanian, 2003], as there is a natural variability in the process due to raw material vari-

ability and due to unsteady environmental conditions. The controller may be expected to perform over a different operating region or meet more stringent performance criteria than originally specified. The controller configuration, parameters, and actions are not only determined by the mathematical models of the process and the controller, but they are also crucially dependent upon whether the assumptions that underlie the mathematical models are still valid, e.g. the sensor can be faulty. When the controller action may be ineffective, the model may have been linearized near the steady-state operating condition, but due to some equipment malfunction the process can be drifting towards a new steady state for a different control configuration, or different set points and gains.

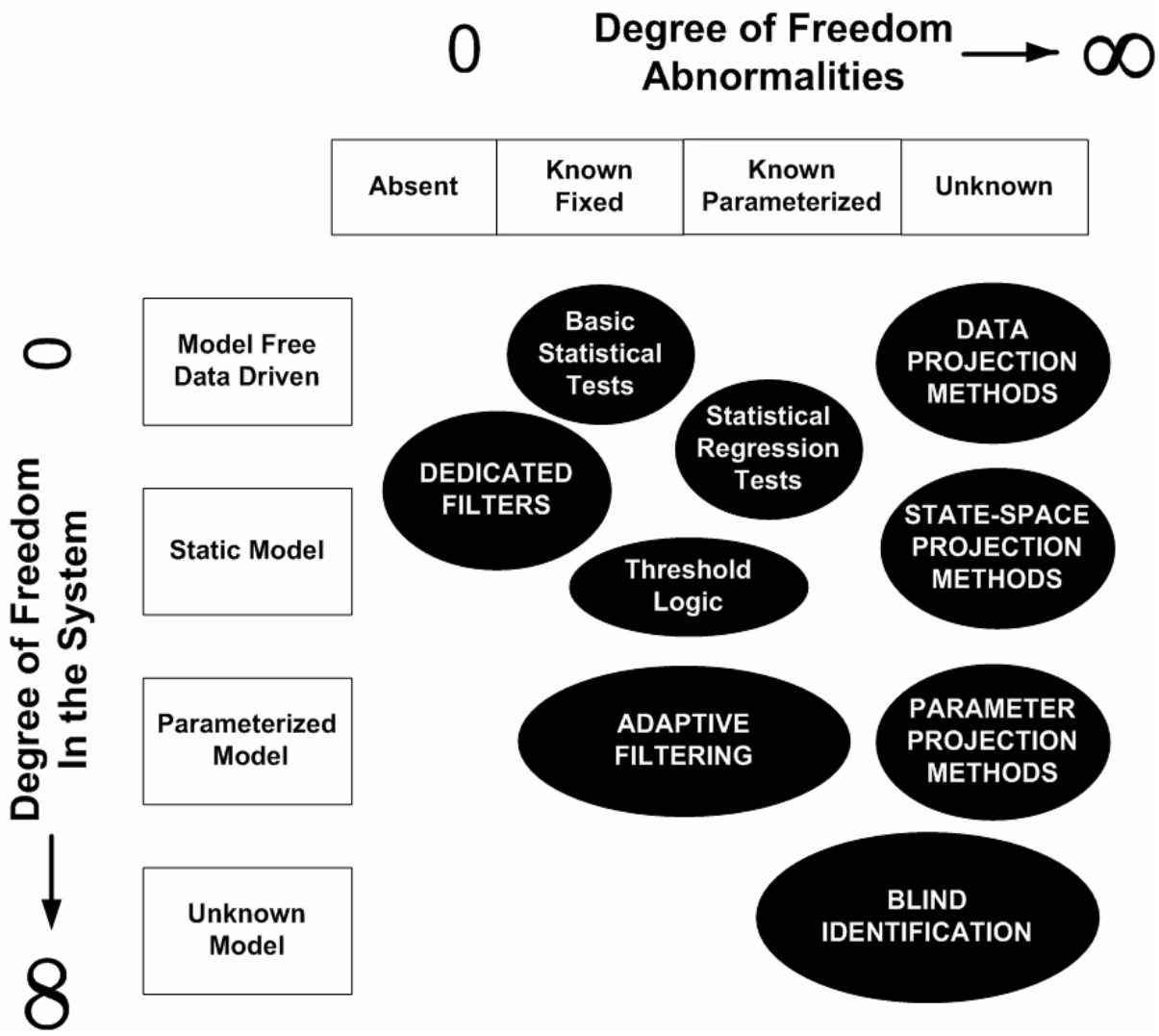

Figure 4.5 : A classification of detection mechanisms, classified by degrees of freedom in models

Hence the modeling of system and abnormality is a key design issue. The disciplines (signal detection, FDI and computational intelligence/neural networks) provide alternative modeling approaches. The a priori knowledge of system and abnormalities leads to a certain preferred mechanism, or design pattern. Variability is one factor that limits the a priori knowledge of the system behavior within it's environment. The less knowledge is available to model the systems and it's disturbances, the more degrees of freedom are required to model the system and it's abnormalities. Hence we can classify the mechanisms by the degrees of freedom of the system model and the model of abnormalities, as is done in figure 4.5.

In this research we focus on controlled systems. Control mechanisms are an integral part of most dynamic systems which we aim to guide towards certain desired behavior. In systems, control is designed purposefully, while in natural processes the dynamics result from collaboration and competition over resources. Key challenges are to deal with variability and to provide the capability of novelty identifiability. At the intersection of variability and novelty detection, the complexity of systems and abnormalities prevents an a priori finite model, and we are stuck with methods that are too coarse for fault prevention.

The strategies based on a priori known finite parameterized models of systems and abnormalities can optimize the detection models and signature computation for sensitivity, robustness and promptness. This optimization is the orthogonalization of projections which distinguish the ideal system behavior from abnormalities, and even distinguish different abnormalities (e.g. sensor, actuator and mode). Such an orthogonalization is only possible if the systems and abnormalities are known a priori. This is also true for the projection methods, recall 4.2.5 and 4.3.3, and note that these methods for detection are based on the idea that abnormalities are in the dual-space of the normal behavior, i.e. the signal or state-sequences visited in case of abnormality are orthogonal to the signal-space for normal behavior. Hence if the basis spanning the normal behavior is known, the ideal projection for abnormality detection can be computed. The use of a system model is essential to relate measurements to states. The actual states and state transitions of the system are compared with those that are desired.

4.6 Summary

We have provided an overview of the conventional detection techniques in different disciplines. The modeling of system and abnormality appears to be the key design issue. The complexity of systems and abnormalities prevents an a priori exact model. In pragmatic engineering approaches a combination of the discussed academic methods is applied to tackle the detection problems. A survey of neural networks in detection illustrates by means of figure 4.3, that neural nets are applied as extensions to a process model based on system-theory, but not as process model by itself; a solution approach that will be explored in chapter 7 and 8.

In the next chapter we analyze the limitations of the existing strategies, discussed here, for detection in locally autonomous distributed systems. Chapter 2 provides metrics for the complexity of systems and models, while chapter 3 discusses the complexity of neural models. This chapter contributes a novel classification of conventional strategies according to complexity of system and abnormality. In chapter 6 we will continue the discussion by exploring the required modeling and estimation capabilities for early detection to deal with the complexity of systems and abnormalities.